Estimating Cavity Tree Abundance Using Nearest Neighbor Imputation Methods for Western Oregon and Washington Forests

Hailemariam Temesgen, Tara M. Barrett and Greg Latta

Cavity trees contribute to diverse forest structure and wildlife habitat. For a given stand, the size and density of cavity trees indicate its diversity, complexity, and suitability for wildlife habitat. Size and density of cavity trees vary with stand age, density, and structure. Using Forest Inventory and Analysis (FIA) data collected in western Oregon and western Washington, we applied correlation analysis and graphical approaches to examine relationships between cavity tree abundance and stand characteristics. Cavity tree abundance was negatively correlated with site index and percent composition of conifers, but positively correlated with stand density, quadratic mean diameter, and percent composition of hardwoods.

Using FIA data, we examined the performance of Most Similar Neighbor (MSN), k nearest neighbor, and weighted MSN imputation with three variable transformations (regular, square root, and logarithmic) and Classification and Regression Tree with MSN imputation to estimate cavity tree abundance from stand attributes. There was a large reduction in mean root mean square error from 20% to 50% reference sets, but very little reduction in using the 80% reference sets, corresponding to the decreases in mean distances. The MSN imputation using square root transformation provided better estimates of cavity tree abundance for western Oregon and western Washington forests. We found that cavity trees were only 0.25 percent of live trees and 13.8 percent of dead trees in the forests of western Oregon and western Washington, thus rarer and more difficult to predict than many other forest attributes. Potential applications of MSN imputation include selecting and modeling wildlife habitat to support forest planning efforts, regional inventories, and evaluation of different management scenarios.

1 Introduction

Forest landscape planning has evolved from a simple harvest scheduling concept to a more detailed analysis involving commodity production as well as habitat preservation/creation and provision of ecosystem services. This change is occurring at all scales and across all ownerships in western Oregon and western Washington, requiring resource management plans to increasingly consider stand, landscape, and forest level attributes. This has led to the need to link and integrate various attributes across scales. To realize this linkage, analysts would like detailed data on every parcel of land of a management area. Usually, interpreted aerial photograph information, such as a forest cover map, is available for every parcel of land, supplemented by detailed ground information on selected parcels. Sample parcels are obtained through two-phase sampling, and provide initial stand condition and tree-lists (a tree-by-tree record).

For habitat modeling, the second phase samples include cavity tree-lists (tree-lists that include attribute information on cavity presence) to be used as input into a habitat model in order to obtain estimates for every parcel of land (e.g., Ohmann et al. 1994, Temesgen and LeMay 2001). Cavity trees are "trees with holes or other structure[s] large enough to shelter animals" (Fan et al. 2003a). In this paper, cavity trees are defined as live or standing dead trees greater than 12.5 cm diameter and 1.37 meters in height, which contain cavities that could be used by wildlife such as birds or mammals.

Cavity trees are an important indicator of wildlife habitat. Despite analysts' desire, complete and detailed information on every parcel of land is rarely available and must be supplied through other methods. Because a forest cover map is usually available for every parcel of land, approaches that estimate or generate the second phase samples from an existing information source (e.g., a forest inventory database) are invaluable.

Most cavities are found in dead standing and recently dead down trees which contribute to forest structure, dynamics and forest succession (Bull et al. 1997, Ganey 1999) and provide habitat for the maintenance or enhancement of wildlife

(e.g., nest cavities, nurseries, etc.) (Ohmann et al. 1994, Ganey 1999, McComb and Lindenmayer 1999, McComb 2007). The size and density of cavity trees provide one indication of the suitability of a forest stand for wildlife habitat (Ohmann et al. 1994, Ganey 1999). Size and density of cavity trees vary with crown closure, understory vegetation, and stand structure (Fan et al. 2003a, Fan et al. 2003b). Because of their importance, forest certification efforts often use wildlife tree retention as a local criteria and indicator to evaluate a management plan of a given area.

In the western United States, the national inventory of public and private forests (the Forest Inventory and Analysis [FIA] inventory) selects approximately one plot per 2400 hectares of land for detailed field measurements, with one tenth of the sample plots measured each year (USDA Forest Service 2005). In some regions, information on cavity occurrence is collected along with other information such as species, diameter, and height of individual trees.

Managers and planners interested in assessing wildlife habitat for their forests would prefer to have tree-related information (such as species composition, forest structure, and snag and cavity occurrence) that is accurate, comprehensive in spatial extent, current, and very detailed. Realistically, such information is far too costly to collect. Instead, the ground-based samples of forest polygons measured in the inventories provide detailed information about trees in the form of tree lists, and then the tree lists are imputed to update the information temporally or extend it spatially to unsampled forest polygons for forest assessment and planning.

Many approaches have been used to generate tree-lists. Broadly, these approaches can be categorized into parametric (e.g., Lindsay et al. 1996), imputation (e.g., Moeur and Stage 1995, Temesgen and LeMay 2001, Temesgen et al. 2003, LeMay and Temesgen 2005), and classification and regression tree (CART) (Fan et al. 2003a, Fan et al. 2004) methods. Although the methods described by these researchers have been used to generate general tree-lists, the methods also have potential for generating cavity tree-lists.

The parametric approach involves fitting a cavity tree diameter distribution for each stand, and then predicting the parameters of a selected

distribution function using stand-level variables. Depending upon how the parameters are predicted, parametric methods have been classified as parameter prediction (Rennolls et al. 1985, Biging et al. 1994, Lindsay et al. 1996), parameter recovery (Bailey and Dell 1973, Bailey 1980, Hyink and Moser 1983), or percentile prediction (Bailey 1980).

Imputation methods can also be used to generate cavity tree-lists from aerial attributes or forest cover maps (Temesgen 2003). Imputation is defined as "replacing missing or non-sampled measurements for any unit in the population with measurements from another unit with similar characteristics" (Ek et al. 1997). Such methods include the nearest neighbor method (Moeur 2000), the most similar neighbor method (Moeur and Stage 1995), the k-nearest neighbor method (Katila and Tomppo 2002, Maltamo and Kangas 1998), geo-statistical estimation (Moeur and Hershey 1998), and tabular imputation models (Ek et al. 1997). Unlike the parametric approach, imputation methods can retain both spatial and attribute structure of the data (Ek et al. 1997, Moeur and Stage 1995); do not restrict the form or shape of the underlying distribution; create projections that will always be within the bounds of biological reality (Moeur and Stage 1995); and can be used to link stand, landscape, and forest level attributes (LeMay and Temesgen 2005).

In the CART method, the model-building process can be seen as a hierarchical refinement of parametric models, similar to forward variable selection in regression analysis. To predict the value of the response variable, the mean value of the response variable in a terminal (leaf) node of the tree is the estimated value. If the response variable is continuous, then a regression tree is generated.

The advantages of regression tree approaches include: they are non-parametric, do not require specification of a functional form (e.g. a general linear model); pre-selection of variables is not needed (a robust stepwise selection method is used); and the same variable can be reused in different parts of a tree as context dependency is automatically recognized. Unlike multiple linear regression or maximum likelihood methods, no single dominant data structure (e.g. normality) is assumed or required; and these methods are robust to the effects of outliers and missing data because surrogate variables can be used for missing values.

Count distributions are useful to describe non-negative integer values such as frequency of snags and cavity trees/ha. Most count data require transformation of the dependent variable. However, little is known about the performance of the imputation methods when variables are transformed. This study utilizes the FIA western Oregon and western Washington inventory and imputes cavity tree lists utilizing two transformations of number of cavity trees/ha (square root and logarithmic transformations) with three different sampling intensities. The goal is to not only assist in determining an appropriate variable transformation method for a given variable, but to provide information of assistance in preparing guidelines for future variable selection and transformation.

Due to prohibitive costs of collecting detailed information over an extensive land base, most stands do not have initial cavity tree-lists or cavity tree abundance for habitat modeling. However, forest cover maps, derived from aerial information or remote sensing data, are often available for every parcel of land in the area of interest. In this manuscript, we examined relationships between cavity tree abundance and stand characteristics, and compared the accuracy of the Most Similar Neighbor (MSN) imputation, the CART and MSN imputation, and variable transformation for their predictive abilities in estimating cavity tree abundance to aid habitat modelling for western Oregon and Washington forests.

2 Methods

2.1 Data

Data for this study were obtained from the Forest Inventory and Analysis (FIA) databases for western Oregon and western Washington. The FIA databases are part of the national inventory of forests for the United States (Roesch and Reams 1999, Czaplewski 1999). A tessellation of hexagons, each approximately 2400 hectares in size, is superimposed across the nation, with one field plot randomly located within each hexagon.

Approximately the same number of plots is measured each year, and each plot has the same probability of selection. Each field plot is composed of four subplots, with each subplot composed of three nested fixed-radius areas used to sample trees of different sizes (Fig. 1). Forested areas that are distinguished by structure, management history, or forest type, are identified as unique polygons (also called condition-classes) on the plot and correspond to stands of at least 0.4047 hectare in size.

In the field, cavity presence was collected by classifying each live or dead tree taller than 1.5 meters and greater than 12.5 cm diameter measured at 1.37 meters above the ground (DBH) into one of three categories: 1) no cavity or den present, 2) cavity or den greater than 15.2 cm diameter is present, and 3) cavity or den less than 15.2 cm diameter is present, no larger cavities are present. A hole in a tree was considered to be a cavity only if, in the field crew's judgment, it could be used by birds or small or large mammals. We assumed cavity tree abundance to be additive from individual trees in a stand, and quantified cavity tree abundance as the number of conifer and hardwood cavity trees per hectare without partitioning it by species.

We selected a number of stand-level variables for examination of relationships to cavity tree abundance. The independent variables included items related to site, forest structure, and general species composition. Collectively, we refer to these items as map label (aerial) variables, because they are typical of attributes available from mapped forest data. For both dependent and independent variables, data were prepared and compiled by mapped polygon (stand). So, for example, stems and volume per hectare are expressed on a stand basis.

Our 2001–2004 FIA data set for Oregon and Washington contained 2285 sample polygons that covered a wide range of ground and aerial variables (Table 1), including a range of diameter at breast height (DBH) values from 12.7 to 200.7 cm, stems per hectare value from 24 to 2296 trees, and predicted volume per ha values from 2 to 4008.7 m^3 per ha. The forest cover map or aerial variables ranged from 6 to 2164 m in elevation, 0 to 150.0 percent slope, and 0.1 to 200.7 cm quadratic mean diameter.
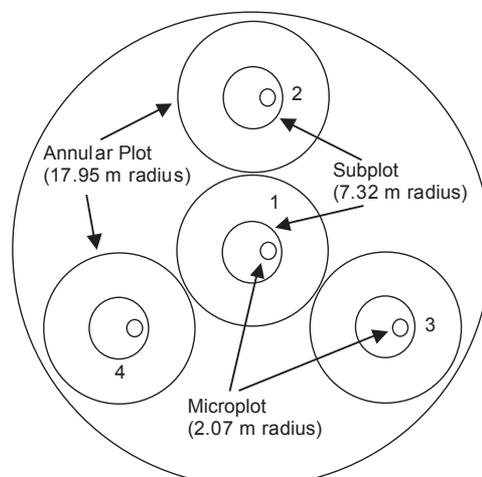
Forest Inventory and Analysis (FIA) plot design

Fig. 1. Each field plot is composed of four subplots, with each subplot composed of three fixed-radius areas used to sample trees of different sizes.

The data set represented many tree species, including Douglas-fir (*Pseudotsuga menziesii*), white fir (*Abies concolor*), Pacific silver fir (*Abies amabilis*), western hemlock (*Tsuga heterophylla*), mountain hemlock (*Tsuga mertensiana*), western red cedar (*Thuja plicata*), red alder (*Alnus oregona* (*rubra*)), and Oregon white oak (*Quercus garryana*).

In this study, stand age represented development stage as a categorical variable, while slope, aspect, and elevation represented biophysical gradients as covariates. Potential site productivity or mean annual increment (MAI) is calculated from the stand's site index, which is itself calculated from age and height of site trees (Hanson et al. 2002).

The Pearson product-moment correlation coefficients were used to measure relationships between cavity tree abundance and stand (and site) attributes. Graphical approaches were used to indicate relationships between cavity tree abundance and stand characteristics in western Oregon and Washington forests. These included the frequency distribution of stands with cavity trees and various stand characteristics.

Table 1. Descriptive statistics for cavity trees, snags, and selected aerial or map variables, n = 2285.

Category	Variable	Minimum	Mean	Median	Maximum	Std
Cavity	No. of conifer trees/ha	0.0	1.8	0.0	264.2	17.7
	No. of hardwood trees/ha	0.0	0.5	0.0	100.5	5.7
	No. of cavity trees/ha	0.0	7.9	0.0	264.2	18.5
	Basal area of cavity trees/ha	0.2	6.4	3.6	61.8	7.7
Snags	Basal area of snags/ha	0.0	6.1	2.4	134.2	9.6
	No. of snags/ha	0.0	55.4	29.7	952.9	80.3
	Average snag diameter (cm)	0.0	28	21.6	28.0	29.5
	Average snag height (m)	0.0	7.8	7.4	52.5	7.0
All	Diameter at breast height (cm)	12.7	32.3	28.8	200.7	16.5
	Height (m)	1.5	18.8	18.8	57.4	7.2
	BA/HA (m^2/ha)	0.2	45.3	40.0	307.0	34.1
	No. of trees/ha	24	497	500	2296	290
	Aspect (deg)	0.0	167.5	170.0	360.0	115.7
	Elevation (m)	0.0	606.1	508.0	2164.0	444.0
	Slope (%)	0.0	35	30.0	150.0	26.4
	Stand age (years)	0.0	87.3	59.0	683.0	96.7
	Mean annual increment (m^3/ha/yr)	0.0	9.3	14.0	41.9	10.9

2.2 Imputation

Using observed and estimated cavity tree abundance and related attributes, the most similar neighbor stand (MSN), variable transformation, and CART approaches were compared for their predictive abilities.

Nearest Neighbor (MSN) Imputation

In this study, the 2285 sample polygons (n) were randomly divided as reference and target polygons. Reference polygons refer to sampled polygons that had both cavity tree attributes and map/aerial attributes, while target polygons refer to un-sampled polygons that only had map/aerial attributes. Reference polygons formed the pool of potential similar neighbors that could be selected to impute cavity tree abundance and related attributes on to target polygons. They were used to develop a similarity function in selecting a neighbour stand in the MSN analysis. The target polygons were assumed to be un-sampled polygons (missing cavity tree abundance data), and were used to validate the accuracy of the MSN approach by comparing the observed cavity tree abundance to the expected cavity tree abundance, obtained by substituting the cavity tree abundance

of the most similar reference polygon using SAS (SAS Institute Inc. 1990).

To link a reference stand to a target stand, the MSN involved two steps. First, canonical correlation between cavity tree abundance (Y set) and selected map/aerial (X set) attributes was used to determine weights. Second, the "most similar" sampled polygon was selected based on the aerial data, weighted by the correlations to the ground data, and linked to a target stand (Moeur and Stage 1995).

For non-sampled polygons (target polygons), cavity tree abundance was estimated from a similar sampled (reference) polygon, by imputing the cavity tree abundance of the reference polygon. A reference stand was selected using weighted squared Euclidean distance, as outlined in Moeur and Stage (1995):

$$D^2_{uj} = (X_u - X_j) \, \Gamma\Lambda^2\Gamma' \, (X_u - X_j)' \qquad (1)$$

Γ is a matrix of standardized canonical coefficients of aerial variables and Λ^2 is a diagonal matrix of squared canonical correlations between map/aerial attributes and cavity tree variables. X_u is a vector of standardized values of the map/aerial variables for the uth target polygon and X_j is a vector of standardized values of the aerial variables for the jth reference polygon. The weighted

Table 2. Variables selected for the most similar neighbour analysis.

Variables of interest (Y set) for cavity tree abundance for habitat modelling	Map label (aerial) and site variables (X set)
Number of conifer cavity trees/ ha Number of hardwood cavity trees/ ha	Percent composition of conifers Stand age (years) Height class midpoint (m) Potential productivity ($m^3ha^{-1}yr^{-1}$) Elevation (m) Aspect (degrees) Slope (%)

squared Euclidean distance was selected over the absolute difference (Maltamo and Kangas 1998) and Euclidean distance (Moeur 2000), as it incorporates the relationships between map/aerial through Λ^2 and cavity tree variables.

To indicate the suitability of a forest stand for wildlife habitat, density of cavity trees were represented by two species groups (i.e., conifers and hardwoods). For each target stand, conifer and hardwood cavity tree abundance and seven map/aerial variables (Table 2) were used to select a reference stand with the smallest weighted squared Euclidean distance. Subsequently, the selected polygon's cavity tree-lists were imputed to the target polygons.

Multiple Neighbor Imputation – KNN and Weighted KNN

Once nearest neighbors are found for the target stand, the imputation of detailed information from the Y set of variables has been based on:

1) Using the information for the nearest neighbor as the substitute (NN; e.g., Moeur and Stage 1995);
2) Using an average of the Y variables over the k nearest neighbors (KNN; e.g., Korhonen and Kangas 1997, Maltamo and Kangas 1998); or
3) Using a weighted average of the k-th nearest neighbors, often based on distance from (or similarity to) the target stand (WKNN; e.g., Maltamo and Kangas 1998).

The choice of how many neighbors to use and what weight to use in calculating the average values is not clear, and is sometimes chosen to meet an objective criterion (e.g., small root mean squared error used by McRoberts et al. 2002

for pixel classification). Tuominen et al. (2003) noted that "The higher the value of k, the more averaging that occurs in the estimates. Thus, the optimal value of k is a trade-off between the accuracy of estimates and the variation retained in the estimates". McRoberts et al. 2002 also noted that as k increases, the biases (average difference between observed and predicted values) rise for extreme values of the variables of interest. Using one neighbor would likely provide the best results if there were a high proportion of stands with full information, since these reference stands would represent the population well. Conversely, if there were a low proportion of stands with full information, using more than one neighbor might give better results, as averaging would provide a wider variety of Y variables. Using too many neighbors might result in less variability in the imputed values than was present in the population because of averaging of values. Also, as noted earlier, values that do not exist in the population may also result via averaging. In this study, for each simulation, imputed values were repeated using the single most similar neighbor (NN), average of three nearest neighbors (KNN), and weighted average of three nearest neighbors (WKNN).

Proportion of Stands with Full Information

The stands with information on all variables form the reference set. If the reference set was based on a simple random selection of stands from the population, a larger proportion should result in better imputation results, as least in matching the X variables, because the reference set would better represent the variability in the population. If the reference set represents the variability present

in the population for the Y set of variables, and the X and Y sets of variables are well correlated, then the imputation should work well. The proportion (or number of observations) needed to obtain a "good" representation of the population will increase with increasing variability of the Y variables in the population, and decrease with increasing correlations between X and Y variables. Also, a greater proportion would be needed as the number of Y variables increases, because it will be difficult to find a match that is similar for all Y variables.

For this simulation, three sampling intensities were selected: 20% (≈ 457 reference polygons), 50% (≈ 1142 reference polygons), and 80% (≈ 1828 reference polygons) of the FIA plots were used as reference polygons, and formed the pool of potential similar neighbours that could be selected to impute cavity tree abundance on to the remaining target polygons. The target polygons were assumed to be un-sampled polygons that only had map/aerial attributes, and were used to validate the accuracy of the MSN under varying sampling intensities by comparing the observed cavity tree abundance to the expected cavity tree abundance, obtained by substituting the cavity tree abundance of the most similar reference polygon using SAS (SAS Institute Inc. 1990). Moeur (2000) indicated that 20% sampling intensity is likely sufficient for estimating stand level variables. LeMay and Temesgen (2005) used simulations to compare the use of different proportions for imputing tree-lists from aerial variables. They noted that there was an improvement in results when the proportion of stands with full information was increased from 20% to 50%, but observed little gain in extending to 80%.

Variable Transformation

Variable transformation affects canonical correlation coefficients. Thus, the nearest neighbour selected as a substitute polygon can change when variables are transformed. Variable selection for NN methods consists of two components: 1) examining variable sets for correlation and 2) selecting significant variables for nearest neighbour imputation (NNI). As a result, selection of significant variables for NNI method is more chal-

lenging than parametric variables selection using stepwise regression or the best selection or other procedures. The transformation of the dependent variables affect the distribution of the Y set variables and the error terms (Ramsey and Schafer 2002). Thus, a set of transformations might help in selecting an ideal nearest neighbour.

In this study, to simplify relationships and to examine the impacts of variable transformation in imputing cavity tree abundance, we transformed the Y set variables using square root and logarithmic transformations. Extensive Monte Carlo simulation studies were conducted to examine the performance of dependent-variable transformations for asymptotic behaviour of the imputation estimates.

The target polygons were assumed to be un-sampled polygons (missing cavity tree abundance data), and were used to validate the accuracy of variable classification and transformation approaches by comparing the observed cavity tree abundance to the expected cavity tree abundance, obtained by substituting the cavity tree abundance of the most similar reference polygon using SAS (SAS Institute Inc. 1990).

CART Partitioning with MSN Imputation

A binary CART (Breiman et al. 1984) approach was used to develop classification rules for estimating cavity tree abundance and to characterize relationships between cavity tree abundance and stand attributes. In fitting the CART models, stand and site variables were used to split the data into increasingly homogenous subsets, using binary recursive algorithms developed in SAS (SAS Institute Inc. 1990). The stand classifying variables included slope and basal area per hectare. The chi-square test was used to split nodes at a significance level of 0.2.

Using the CART method (Breiman et al. 1984), data were partitioned into homogenous groups. Within each node, sample polygons were randomly divided, as reference and target polygons, and then imputation was carried out within each node. The reference polygons were used to develop a similarity function to select a neighbour stand within each node. Within each node, the target polygons were assumed to be un-

sampled polygons (missing cavity tree data), and were used to validate the accuracy of the MSN approach by comparing the observed cavity tree abundance to the expected cavity tree abundance, obtained by substituting the cavity tree abundance of the most similar reference polygon using SAS (SAS Institute Inc. 1990).

2.3 Comparison of Approaches

For each of the 9 combinations (three sample sizes and three transformations), the random separation of the data into target versus reference stands was repeated 200 times. Fit statistics commonly used by other authors are based on comparing observed with estimated values in the simulated target dataset, and in particular, the average difference (often called bias) and root mean squared error (square root of the average squared difference; RMSE) are often calculated.

The average of the imputed Y values will not be an unbiased estimator of the population average (LeMay and Temesgen 2005), even if there is only one variable of interest in the variable-space nearest neighbor imputation. For more than one variable of interest, a small average difference in one variable could be compensated by small average difference in another variable. Also, large negative and positive differences would result in an average difference of zero. The RMSE gives a better indication of the imputation results, because differences are squared prior to averaging.

Moeur and Stage (1995) suggested that the distance metrics could be used to assess the adequacy of results; if distance metrics were high for some stands, then no suitable match was found in the reference set.

To evaluate the results for each simulation, bias (average difference) and RMSE were calculated for each replicate, as follows (after LeMay and Temesgen 2005):

1) Bias for each variable in the Y sets of the target data, as shown for the lth Y variable:

$$bias = \sum_{i=1}^{n} (y_{il} - y_{jl}) \Big/ n \qquad (2)$$

2) RMSE for each variable in the Y sets, as shown for the lth Y variable:

$$RMSE_l = \sqrt{\sum_{i}^{n} (y_{il} - y_{jl})^2 \Big/ n} \qquad (3)$$

where n is the number of stands with missing ground data (target stands) for the replicate. Because a large value for one variable might be compensated by a small value for another variable, these two statistics were also obtained by replicate for all Y variables combined as shown below:

3) Bias for all Y variables combined:

$$bias = \left[\sum_{i=1}^{n} \sum_{j=1}^{all\ Y's} (y_{il} - y_{jl}) \right] \Big/ n \qquad (4)$$

4) RMSE for all Y variables combined:

$$RMSE = \sqrt{\left[\sum_{i=1}^{n} \sum_{j=1}^{all\ Y's} (y_{il} - y_{jl}) \right]^2 \Big/ n} \qquad (5)$$

The mean, minimum, maximum, and range of each of these two statistics were summarized over the 200 sampling replications. In addition, the mean distance was also calculated for each simulation and then averaged over the 200 sampling replications.

For the CART method, weighted RMSE values were calculated and used to compare the predictive abilities of the classification method against other methods considered in this study. Moreover, graphical comparisons of observed (target) and estimated (selected reference) cavity tree abundance were used to examine the predictive abilities of the three approaches. The most similar neighbor stand (MSN), and variable transformation CART approaches were compared for accuracy using observed and estimated cavity tree abundance by size and species composition.
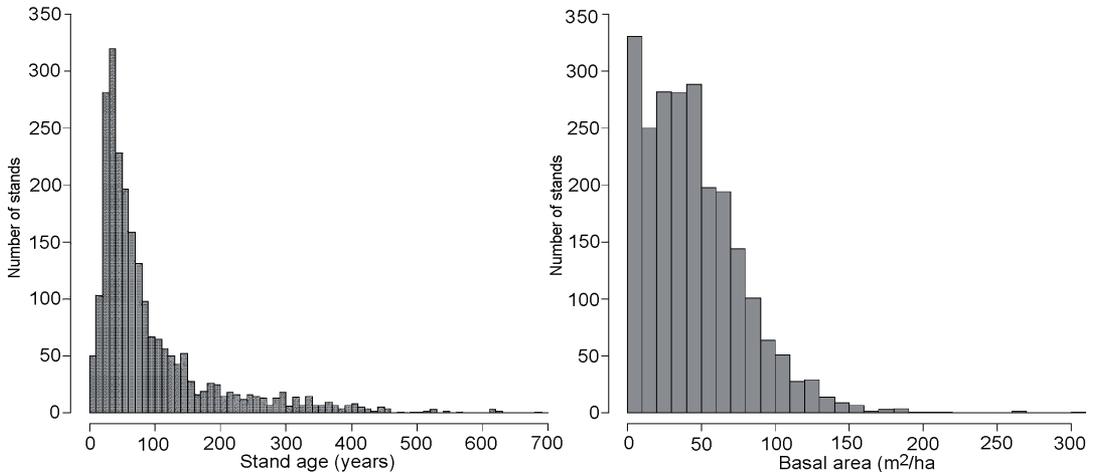
Fig. 2. Frequency distribution of sample polygons by stand age and basal area per ha.

3 Results and Discussion

The 2001–2004 FIA data covered 11.1 million hectares of forestland in western Oregon and western Washington, and indicated that there are an estimated 11.35 million live cavity trees greater than 5.0 cm dbh, which is just 0.25 percent of all live trees. For the same area, there are an estimated 79.8 million dead standing cavity trees, which is 13.8 percent of all dead trees. Because field crew are more likely to miss spotting a cavity in a tree than to see a cavity that is not there, it can be expected that there is some bias toward underestimation in these attributes.

Out of the 2285-forested FIA polygons, 612 (26.8%) polygons did not have any standing dead trees. The remaining 1673 polygons that had at least one snag tree, of which 780 stands had one or more cavity trees (34% of the total polygons had one or more visible cavities). Accordingly, the overall odds ratio of a polygon with no cavity trees (odds of a polygon with no cavity trees/odds of a polygon having cavity trees) in western Oregon and Washington forests was 1.94. In other words, there are 1.94 polygons with no cavity tree for every polygon with one or more cavity trees.

The number of conifer and hardwood cavity trees ranged from 0 to 264 and 0 to 101 trees/ha, while the number of live conifer and hardwood varied from 0 to 2044 and 0 to 1346 trees/ha (Table 1). The distributions of sample FIA poly-
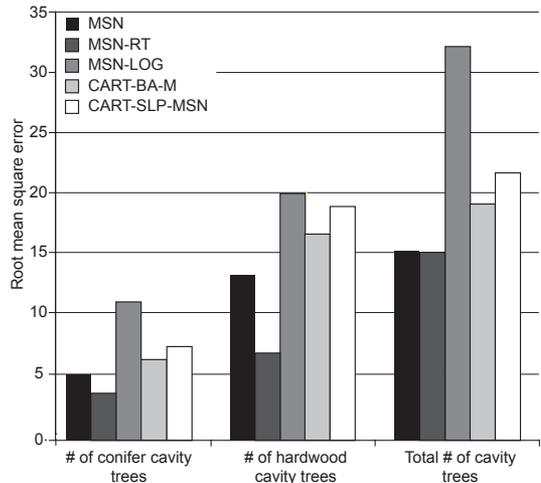
Fig. 3. Performance of imputation methods for estimating cavity tree abundance using 50% sampling intensity. MSN = Most Similar Neighbor; MSR-RT = Most Similar Neighbor with square root transformation; MSN-LOG = Most Similar Neighbor with logarithmic transformation; CART-BA-M = CART classified by basal area with MSN; CART-SLP-MSN = CART classified by slope with MSN.

gons by stand age and basal area/ha depicted a reverse J-shape (Fig. 2). Cavity tree abundance by species and size class graphs showed reasonable matches for some species and poor matches for others (Fig. 3).

Table 3a. Pearson's correlation coefficient and significance between selected variables, n = 2285.

Attribute	Basal area of cavity trees/ha	No. of cavity trees/ha	No. of conifer cavity trees/ha	No. of hardwood cavity trees/ha	Basal area of snags/ha
% conifers	0.09229	0.14303	0.59325	–0.35519	0.20421
	0.0585	<.0001	<.0001	<.0001	<.0001
% hardwoods	–0.07843	–0.04277	–0.33988	0.70775	–0.11859
	0.1081	0.131	<.0001	<.0001	<.0001
Site index (m)	0.01226	–0.0634	–0.06012	0.01932	–0.0932
	0.816	0.0444	0.0566	0.5406	0.0031
Basal area (m^{2}ha^{-1})	0.15363	0.22363	0.44536	0.03804	0.30371
	0.0016	<.0001	<.0001	0.1928	<.0001
Average age (yr)	0.18201	0.1961	0.0633	–0.08195	0.35724
	0.0002	<.0001	0.0254	0.0038	<.0001
Average height (m)	0.20375	0.23506	0.0966	–0.11133	0.35054
	<.0001	<.0001	0.0006	<.0001	<.0001
Slope	–0.0228	0.01772	–0.01859	0.04964	0.03056
	0.6412	0.5322	0.5123	0.0799	0.2813
Aspect	–0.02035	0.03324	0.02758	0.0322	0.03857
	0.6775	0.2412	0.3309	0.2562	0.1738
Elevation (m)	0.04982	0.10466	0.22837	–0.2242	0.20831
	0.3078	0.0002	<.0001	<.0001	<.0001
Volume (m^{3}ha^{-1})	0.15995	0.23858	0.34075	–0.02182	0.34528
	0.001	<.0001	<.0001	0.4411	<.0001
No. of trees per ha	–0.05483	0.10223	0.82681	0.43017	0.07566
	0.2617	0.0003	<.0001	<.0001	0.0075
Mean annual increment (m^{3}ha^{-1}yr^{-1})	–0.01598	–0.02272	0.07316	0.04468	–0.11868
	0.7438	0.4368	0.0122	0.126	<.0001

Table 3b. Pearson's correlation coefficient and significance between selected variables, n = 2285.

Attribute	No. of conifer trees/ha	No. of hardwood trees/ha	No. of cavity trees/ha	Basal area of cavity trees/ha
No. of conifer snags/ha	0.35155	–0.14374	0.43238	0.35028
	<.0001	<.0001	<.0001	<.0001
No. of hardwood snags/ha	–0.08505	0.28823	0.06080	–0.03931
	0.0026	<.0001	0.0317	0.4211
Basal area of snags/ha	0.14149	–0.09414	0.58930	0.83104
	<.0001	0.0009	<.0001	<.0001
No. of snags/ha	0.27228	0.01244	0.41564	0.32407
	<.0001	0.6606	<.0001	<.0001
Average snag diameter (cm)	0.02842	–0.02880	0.20697	0.29475
	0.3158	0.3093	<.0001	<.0001
Average snag height (m)	0.17822	0.01439	0.20460	0.14428
	<.0001	0.6114	<.0001	0.003

Table 4. Descriptive statistics for selected variables by stand development stages, n = 2285.

Variable	Younger than 40 years, n = 753				40 to 90 years, n = 812				Older than 90 years, n = 720			
	Min	Median	Max	Std	Min	Median	Max	Std	Min	Median	Max	Std
Average diameter (cm)	12.7	21.5	200.7	17.9	13.4	29.1	137.2	11.3	17.8	37.4	152.0	16.3
Average height (m)	1.5	14.0	38.9	5.6	4.1	19.9	45.0	6.6	4.6	20.7	57.4	7.4
Snag diameter (cm)	0.0	14.1	228.6	32.7	0.0	20.2	190.5	22.4	0.0	34.7	191.5	28.9
No. of cavity trees	0.0	0.0	83.1	7.4	0.0	0.0	264.2	20.0	0.0	2.5	191.4	22.8
No. of snags	0.0	2.7	757.7	52.4	0.0	35.1	952.9	90.5	0.0	60.1	602.3	81.2
No conifer cavity trees	0.0	0.0	83.1	6.9	0.0	0.0	264.2	18.6	0.0	2.5	191.4	22.3
No. hardwood cavity trees	0.0	0.0	45.1	2.8	0.0	0.0	100.5	7.3	0.0	0.0	74.9	5.7
Snag basal area (m^2/ha)	0.0	0.3	51.2	5.5	0.0	2.2	134.2	8.3	0.0	8.4	71.4	11.6
Cavity tree basal area (m^2/ha)	0.2	2.7	29.1	4.9	0.2	2.7	61.8	7.2	0.2	5.4	59.8	8.4

3.1 Relationships

Cavity tree abundance was negatively correlated with site index and percent composition of conifers, but positively correlated with stand density, quadratic mean diameter, and percent composition of hardwoods. Thirty-four percent of the forested FIA plots had a least one cavity tree, and 29% of the sampled plots had more than 4.1 m^2/ha basal area of cavity trees. No significant correlation was observed between number of hardwood cavity trees and basal area/ha and mean annual increment (Table 3). The number of cavity trees (cavity tree abundance) was strongly related to snag density, snag basal area, or other snag attributes (Table 4). Cavities were more often detected on bigger and taller tree snags that had larger DBH and height. This is consistent with other studies (Screiber and de Calesta 1992).

There was significant correlations between basal area of cavity trees per ha and stand age, average tree height, and quadratic mean diameter. No significant correlations were observed between basal area of cavity trees or cavity tree abundance and slope or aspect of stands (Table 3). Low correlations between cavity tree abundance and most of the predictor variables considered in this study indicate that either predictor variables are not useful or cavity tree abundance is highly variable.

Cavity tree abundance generally increased with stand development stages. In early developmental stages (S40), the proportion of stands with snags was low, and stands contained the fewest number of cavity trees; 17.3% of these plots had at least on cavity tree, while at late developmental stages (>90 years, S90) 55% of the plots had at least one cavity tree (Fig. 2).

Within each stand development stage, cavity tree abundance was found to be highly variable, and the average snag size increased with stand development stage (14.1, 20.2, and 34.7 cm for the three stand development stages). The basal area of snags ranged from 0 to 51.2 m^2/ha, 0.2 to 134.2 m^2/ha, and 0 to 71.4 m^2/ha, while the basal area of cavity trees ranged from 0.2 to 29.1 m^2/ha, 0.2 to 61.8 m^2/ha, and 0.2 to 59.8 m^2/ha, respectively, for the three stand development stages (Table 4).

In mid-developmental stages (40 to 90 years, S60), the proportion of the snags was high, and the ranges of the number of snags and cavity trees were wider than those observed for mature developmental stages (S90). Proportion of stands with at least one cavity tree increased with stand age. Cavity trees were most abundant and tallest in the mature development stage. Number of cavity trees varied more in S90 than in S40, as shown by a standard deviation that was 3.2 times higher. This variation would have a large influence on wildlife habitat selection and management. High variability of cavity tree abundance in the mature development stage is also reported by Ohmann et al. (1994) and Fan et al. (2005).

The number of snags differed significantly among the three age classes, steadily increasing with stand age (Table 4). Stands older than 90 years had 80% higher number of snags than stands 40 to 90 years old. The number of snags and cavity trees of each stand age class were significantly different from one another. These differences may be due to increased natural mortality over time and stand deterioration.

Table 5. Minimum, mean, and maximum root mean square (RMSE) of three variable transformation approaches used to estimate cavity tree abundance.

Attribute	KNN			MSN			WKNN		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
ORIGINAL UNIT									
20%									
No. of conifer cavity trees	9.3	8.4	10.1	8.8	7.6	10.4	9.1	6.4	11.1
No. of hardwood cavity trees	21.8	20.7	23.8	21.7	20	23.9	21.3	14.8	23.9
Total number cavity trees	25.2	24.3	27.2	25	23	27	24.6	17.3	27.2
50%									
No. of conifer cavity trees	5.9	5.4	6.3	5.5	5	5.9	5.9	5.4	6.3
No. of hardwood cavity trees	13.7	13.1	14.2	13.7	13	14.4	13.7	13.1	14.2
Total number cavity trees	15.8	15.3	16.4	15.7	15.1	16.3	15.9	15.3	16.4
80%									
No. of conifer cavity trees	4.7	4.4	4.9	4.4	4.1	4.6	4.3	4	4.5
No. of hardwood cavity trees	10.8	10.4	11.2	10.8	10.4	11.2	9.9	9.4	10.2
Total number cavity trees	12.5	12.2	12.9	12.4	12.1	12.7	11.4	11	11.8
SQUARE ROOT TRANSFORMATION									
20%									
No. of conifer cavity trees	5.7	4.7	6.6	6	4.6	8	5.5	4.8	6.3
No. of hardwood cavity trees	10.6	9.4	11.8	11.3	9.4	13.2	8.7	7.9	9.5
Total number cavity trees	25.2	24.3	27.2	24.8	22	27	23.1	22.3	24.8
50%									
No. of conifer cavity trees	3.7	3.2	4.2	3.9	3.1	4.6	3.4	3.1	3.6
No. of hardwood cavity trees	6.8	6.5	7.2	7.3	6.7	8	5.7	5.3	5.9
Total number cavity trees	15.9	15.3	16.4	15.7	15	16.3	14.8	14.3	15.2
80%									
No. of conifer cavity trees	3.1	2.8	3.4	3.2	2.8	3.8	2.8	2.4	3.3
No. of hardwood cavity trees	5.6	5.3	6	6	5.5	7	4.7	4.4	5.4
Total number cavity trees	12.5	12.2	12.9	12.4	12.1	12.7	11.6	11.4	11.9
LOGARITHMIC TRANSFORMATION									
20%									
No. of conifer cavity trees	11.5	9.9	13.5	17.8	13.6	24.4	16.3	13.7	19.1
No. of hardwood cavity trees	20.5	18.7	22.4	31.9	27.3	35.7	27.5	24.7	30.5
Total number cavity trees	42.7	40.8	44.4	67.6	59.9	73.5	43.4	41.6	48.4
50%									
No. of conifer cavity trees	9.3	8	11.5	11.5	9.9	13.5	10.1	9.2	11.1
No. of hardwood cavity trees	17	15.2	19	20.5	18.7	22.4	17.4	16.3	19
Total number cavity trees	33.8	32.9	34.7	42.7	40.8	44.4	33.3	31.6	34.3
80%									
No. of conifer cavity trees	8.5	7.8	9.3	9.4	8	11.5	8.5	7.8	9.9
No. of hardwood cavity trees	15.5	14.4	16.7	17	15.2	19	14.4	12.9	16
Total number cavity trees	29.5	28.6	30.5	33.8	32.9	34.7	29.5	28.6	30.5

3.2 Predictions

For the seven aerial/map label variables and two ground variables, the canonical correlations were very strong with 0.58, 0.86, and 0.95 for the first, second, and third canonical variates, respectively. The first four variates had canonical correlations greater than 0.99. The number of conifer cavity trees had high linear canonical correlation coefficients for three of these four variates. Other variables that weighted highly were number of hardwood trees per hectare and basal area per hectare. Other variables such as elevation had lower coefficients on the first six covariates, and

Table 6. Minimum, mean, and maximum root mean square by two classifying variables

CART & Attribute	Mean	KNN Min	Max	Mean	MSN Min	Max	Mean	WKNN Min	Max
SLOPE									
20%									
No. of conifer cavity trees	13.1	9.9	17.1	30.6	25.8	35.1	10.8	9.1	13.3
No. of hardwood cavity trees	30.7	26.0	34.6	12.3	9.0	16.4	25.2	22.4	27.5
Total number cavity trees	35.5	31.2	39.7	35.2	30.9	39.5	29.2	27.0	31.5
50%									
No. of conifer cavity trees	8.3	7.2	9.5	17.1	15.4	18.7	8.5	7.6	9.4
No. of hardwood cavity trees	19.4	17.7	20.9	6.7	5.9	7.7	20.0	19.0	21.3
Total number cavity trees	22.4	20.9	23.8	19.6	17.9	21.0	23.2	22.1	24.5
80%									
No. of conifer cavity trees	6.6	6.1	7.1	15.3	14.7	15.9	6.6	6.1	7.1
No. of hardwood cavity trees	15.3	14.7	15.9	6.2	5.7	6.6	15.3	14.7	15.9
Total number cavity trees	17.7	17.1	18.4	17.6	17.0	18.2	17.7	17.1	18.4
BA									
20%									
No. of conifer cavity trees	13.0	9.6	17.6	18.3	16.2	20.7	13.0	9.6	17.6
No. of hardwood cavity trees	30.6	25.8	35.1	13.4	11.7	16.1	30.6	25.8	35.1
Total number cavity trees	35.5	31.2	39.8	25.8	23.9	28.1	35.5	31.2	39.8
50%									
No. of conifer cavity trees	8.3	7.2	9.5	19.4	17.7	20.9	8.3	7.2	9.5
No. of hardwood cavity trees	19.4	17.7	20.9	7.8	6.7	8.9	19.4	17.7	20.9
Total number cavity trees	22.4	20.9	23.8	22.2	20.7	23.6	22.4	20.9	23.8
80%									
No. of conifer cavity trees	6.6	6.1	7.1	15.3	14.7	15.9	6.6	6.1	7.1
No. of hardwood cavity trees	15.3	14.7	15.9	6.2	5.7	6.6	15.3	14.7	15.9
Total number cavity trees	17.7	17.1	18.4	17.5	17.0	18.2	17.7	17.1	18.4

therefore, were weighted less in the Most Similar Neighbour distance measurement.

Bias and RMSE for hardwood cavity trees were consistently higher than bias and RMSE obtained for conifer cavity trees. The square transformation consistently resulted in lower RMSE values than the original unit, while logarithmic transformation resulted in the highest RMSE values (Table 5). For the MSN approach, there was a reasonable match in aerial variables between the target and reference polygons. As expected, weighting the reference polygons by squared Euclidean distance consistently resulted in the lowest RMSE values for all ground variables (Table 6).

Unlike the results reported by Maltamo and Kangas (1998) for more simple Scots pine stands, the MSN resulted in lower biases (absolute values) in cavity tree abundance by species class than other distances examined in this study.

3.3 CART Partitioning and MSN Imputation

CART models were fitted by splitting the data into homogenous subsets or groups and establishing a set of classification rules for predicting the status (alive or dead) of trees. The number of cavity trees was then estimated for each subset. In this study, stand variables (i.e., slope and basal area per hectare) were important in estimating cavity tree abundance and characterizing relationships between cavity tree abundance and stand attributes. The CART method indicated that basal area of 40 m^2/ha and slope of 30% were thresholds for distinguishing cavity trees from non-cavity trees at different stand development stages. Using the CART method, the stands were thus classified into two basal area classes ($\leq$40 m^2/ha and >40 m^2/ha) and two slope classes ($\leq$30% and >30%).

Most classifications obtained using the CART methods were surprisingly simple with usually no more than two splits (levels of data grouping). For example, the selected CART models for proportion of stands with at least one cavity tree showed that the first split used slope at 30% as a threshold.

When slope and basal area per ha were considered, the CART method indicated clear distinction or clear classifying rules for estimating cavity tree abundance. Given this clear separation of cavity tree abundance, imputation within each node or group was expected to result in more precise cavity tree abundance estimates than imputing from the entire data set. However, CART partitioning and imputing (MSN) within each node resulted in marginally worse estimation of the number of cavity trees/ha than imputing from the entire sample. This can be ascribed to the conversion of continuous variables into discrete variables.

3.4 Variable Transformations

Distances cannot be compared across the three variable transformation methods. However, comparisons across the three variable transformation methods (original unit, square root, and logarithmic) and proportions of stands with full information were made for each variable transformation method.

Distances were lowest for square root transformation, followed by the original unit and then by logarithmic transformation. The short distances might have resulted in better matches and lower squared Euclidean distances using square root transformation and original unit. Conversely, the distances using logarithmic transformation were very large and this might have resulted in poor matches. The possible reasons for this improvement might be ascribed to that the assumption about linear correlation was better met with the square root transformation, followed by the original unit and then by logarithmic transformation. The square root transformation (power = 0.5) might have resulted in better representation in multivariate space and more accurate NN predictions because it moves the sparse tails closer to the denser center of the multivariate distribution

than the original unit untransformed (power = 1) feature space and the logarithmic (power = 0) transformation.

Distances decreased with the increasing proportion of stands with full information, with the greatest gain from 20% to 50%, and less gain from 50% to 80%. This result is similar to that noted by LeMay and Temesgen (2005) for imputing tree-lists. As a result, 50% (≈ 1142 reference polygons) of the stands with full information would be preferred for imputing the number of cavity trees in western Oregon and western Washington forests.

3.5 Average Differences

For the original unit and square root transformations, the bias (average difference) for the combined Y variables averaged over the 200 sampling replications (mean bias) was close to zero for the 80% reference sets, while the bias was higher for 20% reference sets. For the 20% and 80% reference sets, original unit and square root transformations gave lower (absolute value) mean biases using all three methods, indicating an improvement via weighting the X variables using correlations with the Y variables. For both variable sets (slope and BA), there was no noticeable reduction in the absolute value of mean bias in using the KNN or WKNN methods over using a single neighbor (NN), nor in using the 50% over the 20% reference sets.

For the variables of interest, the number of cavity trees per ha (Y variable), the mean biases were not close to zero. In their simulations, Moeur and Stage (1995) showed percent biases of –4.0 to 0.6% using Eq. 2 and NN to impute volume variables from land classification variables. For imputing diameter distributions, Maltamo and Kangas (1998) also obtained non-zero biases. Mean biases in the Y variables tended to increase with an increase in the proportion represented in the reference sets. For variable set 1 (slope), mean biases were nearly zero for the 20% reference sets, for all methods. Also, mean biases tended to be lower using square root transformation, as might be expected, because the canonical correlations between the X and Y variables are used to weight the distance metric. Using square root

transformation, some reduction in mean bias was obtained through using the KNN and WKNN approaches, over the NN approach. For variable set 2 (BA), mean biases were again lower using square root transformation, and some reduction in mean bias occurred using the KNN and WKNN approaches.

3.6 Mean, Minimum, and Maximum RMSEs

The RMSE for the Y variables, generally there were some reductions in mean RMSE values, although this varied with variable transformation and weighting alternatives (Tables 5 and 6). Generally, lower mean RMSE values were obtained using KNN and WKNN over NN for square root transformation, with no large difference between KNN and WKNN. For variable set 1 (slope CART), square root transformation was slightly better, particularly for KNN and WKNN. For variable set 2 (BA CART), square root transformation performed slightly better than original unit. Because the stems per ha values are very large compared to the basal area per ha values, the mean RMSE total number of cavity trees per ha better reflects the observed number of cavity trees per ha.

Generally, the minimum of the 200 RMSE values for each Y variable decreased with increasing proportions of stands with full information for all simulations, indicating improvements with an increase in the number of reference stands (Tables 5 and 6). However, the maximum of the 200 RMSE values decreased for 50% over 20% of stands with full information, but then increased for the 80%. Because the RMSE is calculated using target stands only, a large squared difference between observed and estimated values for one target stand would have more impact on the RMSE for the 80% proportion. For example, given a hypothetical number of cavity trees per ha values of 30, 30, 35, 35, 40, 40, 38, 38, 80, and 80, eight stands would be selected as the reference set and two stands as the target stands using the 80% proportion. RMSE values would be small, except when the two stands of 80 cavity tree per ha were selected as the target stands in the simulation. Using 50%, the large maximum RMSE value obtained with the 80% proportion

would not occur, because the RMSE values would be averaged over five stands. In application, this would translate into a situation where the number of target stands is few, and these differ greatly from the reference stands. As noted by Moeur and Stage (1995), the distance metric should indicate this problem. In this case, approaches other than nearest neighbor methods might give better results. While the minimum RMSE may be useful in comparing methods, the maximum RMSE value must, therefore, be interpreted with caution.

Slightly smaller minimum and maximum RMSE values were consistently obtained using KNN and WKNN over NN. The smallest minimum RMSE values were obtained using square root transformation, WKNN, 80% of stands with full information. Using the average of three stands to impute stems per ha and basal area per ha reduced the possibility of a poor match when the target stands were quite different than the remainder of the stands.

3.7 Overall Discussion

The size and density of cavity trees indicate the suitability of a forest stand for wildlife habitat (Ohmann et al. 1994, Ganey 1999) and are important in managing habitat for some wildlife species. For example, Thomas et al. (1979) suggested that American kestrels (*Falco sparverius*) and northern flickers (*Colaptes aurtus*) require nest trees with a minimum diameter of 30 cm in the Pacific Northwest region of North America. Forest managers need insight into sampling intensities and methods that can incorporate cavity tree estimation into their forest plans.

Compared to parametric approach, such as regression analysis, NNI approaches do not require distributional assumptions. For cavity tree abundance, NNI provides the advantage of predicting total number of cavity trees by species group in one step. Imputation also allows spatial modeling of cavity tree densities, which can be important to assessment of potential wildlife home ranges. The approaches examined in this study can estimate cavity abundance from map labels or aerial attributes and can used in other applications such as landscape modeling or 3-D

visualization (McGaughey 1998, McCarter et al. 1998).

Imputation methods can fill in cavity tree abundance by species and tree size classes for non-sampled polygons or panels. The results from this study indicated that generally MSN performed better at this than KNN and WKNN. Using CART to separate observations by basal area or slope classes may not be optimal. However, the sampled polygon areas used in this study were small (≤0.4 hectares) and results may differ when larger areas are used for sampling.

The approaches used in this study can be tailored to meet unique sustainable resource management by selecting or emphasizing on given tree, stand and landscape level variables. For example, if the objective of a management intervention is timber, emphasis can be given to attributes such as total basal area and stand volume. If an objective of a management intervention focuses on tree-size dependent wildlife species, then emphasis can be given to the number of wildlife and large-diameter trees.

It is also important for forest managers to understand the limitations of different modeling techniques. Our study found that cavity trees were only 0.25 percent of live trees and 13.8 percent of dead trees in the forests of western Oregon and western Washington, thus rarer and more difficult to predict than many other forest attributes. This research found that while significant relationships do exist between stand variables and cavity tree presence, correlations were low, an indication of the high variability of cavity trees and low predictive ability from ancillary variables. While it is becoming increasingly common for managers to use nearest neighbor imputation to link forest inventories with forest planning, cavity tree abundance may be one of the more difficult forest attributes to model well.

4 Conclusions

Increased knowledge of cavity tree abundance on every parcel of land provides flexibility for forest management systems and provides tools for resource managers to analyze information from limited data sources, integrate available resource data, and visualize and model ecosystems functioning. The approaches used to estimate cavity-tree abundance offer several potential advantages over classical estimation procedures for sustainable resource management, as some of these methods can preserve spatial and attributes of natural resource data, including their relationships and natural variability. When these methods are employed, projections will always be within the bounds of biological reality, and scenario analysis can account for the inherent correlations between resource attributes. However, like most other methods, the accuracy of the CART and imputation methods is dependent on the representativeness of the sample and the similarity of the target and reference polygons. Further studies to relate cavity tree abundance with cavity nesting birds, flying squirrels, and snag conditions including decay classes are warranted.

Acknowledgments

References

Bailey, R.L. 1980. Individual tree growth derived from diameter distribution model. Forest Science 26: 626–632.

— & Dell, T.R. 1973. Quantifying diameter distribution with the Weibull function. Forest Science 19: 97–104.

Biging, G.S., Robards, T.A., Turnblom, E.C. & Van Deusen, P.C. 1994. The predictive models and procedures used in the forest stand generator. STAG. Hilgardia 61(1): 1–36.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. Classification and regression trees. Wadsworth and Brooks Advanced Books and Software, Monterey, CA. 358 p.

Bull, E.L., Parks, C.G. & Torgersen, T.R. 1997. Trees and logs important to wildlife in the interior Columbia River Basin. USDA Forest Service, General Technical Report PNW-GTR-391. Pacific Northwest Research Station, Portland, Oregon.

Czaplewski, R.L. 1999. Forest survey sampling designs: a history. Journal of Forestry 97: 4–10.

Ek, A.R., Robinson, A.P., Radtke, P.J. & Walters, D.K. 1997. Development and testing of regeneration imputation models for forests in Minnesota. Forest Ecology and Management 94: 129–140.

Fan, Z., Larson, D.R., Shifley, S.R. & Thompson, F.R. 2003a. Estimating cavity tree abundance by stand age and basal area, Missouri, USA. Forest Ecology and Management 179: 231–242.

—, Shifley, S.R., Spetich, M.A. & Thompson, F.R. 2003b. Distribution of cavity trees in midwestern old-growth and second growth forests. Canadian Journal of Forest Research 33: 1481–1494.

—, Shifley, S.R., Thompson, F.R. & Larsen, D.R. 2004. Simulated cavity tree dynamics under alternative timber harvest regimes. Forest Ecology and Management 193: 399–412.

—, Shifley, S.R., Spetich, M.A., Thompson, F.R. & Larsen, D.R. 2005. Abundance and size distribution of cavity trees in second-growth and old-growth central hardwood forests. Northern Journal of Applied Forestry 22(3): 162–169.

Ganey, J.L. 1999. Snag density and composition of snag populations on two national forests in northern Arizona. Forest Ecology and Management 117: 169–178.

Hanson, E.J., Azuma, D.L. & Hiserote, B.A. 2002.

Site index equations and mean annual increment equations for Pacific Northwest Research Station forest inventory and analysis inventories 1985–2001. USDA Forest Service, Research Note PNW-RN-533. 24 p.

Hyink, D.M. & Moser, J.W. Jr. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science 29: 85–95.

Katila, M. & Tomppo, E. 2002. Stratification by ancillary data in multisource forest inventories employing k-nearest-neighbour estimation. Canadian Journal of Forest Research 32: 1548–1561.

Korhonen, K.T. & Kangas, A. 1997. Application of nearest-neighbor regression for generalizing sample tree information. Scandinavian Journal of Forest Research 12: 97–101.

LeMay, V. & Temesgen, H. 2005. Comparison of nearest neighbour methods for estimating basal area and stems per ha using aerial auxiliary variables. Forest Science 51(2): 109–119.

Lindsay, S.R., Wood, G.R. & Woollons, R.C. 1996. Stand table modeling through the Weibull distribution and usage of skewness information. Forest Ecology and Management 81: 19–23.

Maltamo, M. & Kangas, A. 1998. Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. Canadian Journal of Forest Research 28: 1107–1115.

McCarter, J.B., Wilson, J.S., Baker, P.J., Moffett, J.& Oliver, C.D. 1998. Landscape management through integration of existing tools and emerging technologies. Journal of Forestry 96(6): 17–23.

McGaughey, R.J. 1998. Techniques for visualizing the appearance of forestry operations. Journal of Forestry 96(6): 9–14.

McComb, B.C. 2007. Wildlife habitat management: concepts and applications in forestry. Taylor and Francis Publishers, CRC Press, Boca Raton, FL. 319 p.

McComb, W.C. & Lindenmayer, D. 1999. Dying, dead, and down trees. In: Hunter, M.L. Jr. (ed.). Maintaining biodiversity in forest ecosystems. Cambridge University Press, Cambridge, England. 695 p.

McRoberts, R.E., Nelson, M.D. & Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Remote Sensing of Environment 82: 457–468.

Moeur, M. 2000. Extending stand exam data with most similar neighbor inference. In: Proceedings of the Society of American Foresters National Convention. Sept. 11–15, 1999, Portland, Oregon. p. 99–107.

— & Hershey, R. 1998. Preserving spatial and attribute correlation in the interpolation of forest inventory data. In: International symposium on spatial accuracy assessment in natural resources and environmental sciences. May 20–22, 1998, Quebec City, Quebec, Canada.

— & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Science 41: 337–359.

Ohmann, J.L., McComb, W.C. & Zumrawi, A. 1994. Snag abundance for primary cavity-nesting birds on nonfederal forestlands in Oregon and Washington. Wildlife Society Bulletin 22: 607–620.

Ramsey, F.L. & Schafer, D.W. 2002. The statistical sleuth: a course in methods of data analysis. 2nd ed. Duxbury, New York. 741 p.

Rennolls, K., Geary, D.N. & Rollinson, T.J.D. 1985. Characterizing diameter distributions by the use of Weibull distribution. Forestry 58: 57–66.

Roesch, F.A. & Reams, G.A. 1999. Analytical alternatives for an annual inventory system. Journal of Forestry 97: 44–48.

SAS Institute Inc. 1990. SAS/STAT User's guide. Version 6, fourth ed., vol. 2. SAS Institute Inc., Cary, NC. 848 p.

Screiber, B. & de Calesta, D.S. 1992. The relationship between cavity-nesting birds and snags on clearcuts in western Oregon. Forest Ecology and Management 50: 299–316.

Temesgen, H. 2003. Estimating tree-lists from aerial information: a comparison of a parametric and most similar neighbour approaches. Scandinavian Journal of Forest Research 18: 279–288.

— & LeMay, V.M. 2001. Generating tree-lists from aerial attributes for sustainable resource management. In: Proceedings of the "Forest modeling for ecosystem management, forest certification, and sustainable management" conference. Aug. 12–17, 2001, Vancouver, BC, Canada. p. 409–421.

— , LeMay, V.M, Marshall, P.L. & Froese, K. 2003. Imputing tree-lists from aerial attributes for complex stands of British Columbia. Forest Ecology and Management 177: 277–285.

Thomas, J.W., Anderson, R.G., Maser, C. & Bull, E.L. 1979. In: Thomas, J.W. (ed.). Wildlife habitats in managed forests – the Blue Mountains of Oregon and Washington. USDA Forest Service, Agriculture Handbook 553: 60–77.

Tuominen, S., Fish, S. & Poso, S. 2003. Combining remote sensing, data from earlier inventories, and geostatistical interpolation in multisources forest inventory. Canadian Journal of Forest Research 33: 624–634.

USDA Forest Service. 2005. Forest Inventory and Analysis. http://www.fs.fed.us/pnw/fia/ [Accessed 25 October 2005].

Total of 39 references